Distortion Representation of Forecast Errors for Model Skill Assessment and Objective Analysis[1] Technical Report[2]

Ross N. Hoffman, Thomas Nehrkorn and Christopher Grassotti[3]
Atmospheric and Environmental Research, Inc.[4]

Revision 1.8[5]

August 8, 1996

[1]Supported by NASA contract NAS5-32953. AER, Inc. intends to retain patent rights to certain aspects of the algorithms described herein under FAR 52.227-11.

[2]To be submitted to Goddard Space Flight Center (NASA/GSFC), Greenbelt, MD 20771.

[4]Atmospheric and Environmental Research (AER), Inc., 840 Memorial Drive, Cambridge, MA 02139–3794. Phone: +1 617 547 6207. Fax: +1 617 661 6479. Net: http://www.aer.com/.

[5]AER document version control: P584, tr,v 1.8 1996/08/08 20:11:13 rnh Exp. Formatting version start.doc,v 1.2 1996/01/25 21:39:21 rnh Exp rnh.

1 Introduction

We proposed a novel characterization of errors for numerical weather predictions. In its simplest form we decompose the error into a part attributable to phase errors and a remainder. The phase error is represented in the same fashion as a velocity field and will be required to vary slowly and smoothly with position. A general distortion representation allows for the displacement and amplification or bias correction of forecast anomalies.

Characterizing and decomposing forecast error in this way has two important applications, which we term the assessment application and the objective analysis application. For the assessment application, our approach will result in new objective measures of forecast skill which are more in line with subjective measures of forecast skill and which will be useful in validating models and diagnosing their shortcomings. With regard to the objective analysis application, meteorological analysis schemes balance forecast error and observational error to obtain an optimal analysis. Presently, representations of the error covariance matrix used to measure the forecast error are severely limited. For the objective analysis application our approach will improve analyses by providing a more realistic measure of the forecast error. We expect, *a priori*, that our approach should greatly improve the utility of remotely sensed data which have relatively high horizontal resolution, but which are indirectly related to the conventional atmospheric variables.

In this project we are initially focusing on the assessment application, restricted to a realistic but univariate 2–dimensional situation. Specifically we study the forecast errors of the 500 hPa geopotential height field for forecasts of the short and medium range. Since the forecasts will be generated by the GEOS (Goddard Earth Observing System) data assimilation system with and without ERS 1 scatterometer data, these preliminary studies will serve several purposes. They will (1) provide a testbed for the use of the distortion representation of forecast errors, (2) act as one means of validating the GEOS data assimilation system and (3) help to describe the impact of the ERS 1 scatterometer data.

2 Data

Forecasts and verifying analyses made with the GEOS data assimilation and forecast system (Schubert *et al.* 1993) are used here. The particular experiments studied here are described by Atlas *et al.* (1995). These forecasts form the basis of a current collaborative study of the impact of ERS 1 scatterometer data on numerical weather prediction. Subjective assessment of these forecasts is already underway. The period of study is March, 1993. The forecast model and data assimilation system used in these experiments are identical to the GEOS-1 system described by Schubert *et al.*, except for some minor bug fixes and the modifications necessary to utilize surface wind vectors. Thus the control forecasts in the impact study are standard GEOS forecasts. In addition to the CONTROL experiment, several using different types of scatterometer wind information are available. Experiments were conducted using $4 \times 5°$ and $2 \times 2.5°$ versions of the GEOS system. For our initial prototyping and sensitivity studies we are using only the 500 hPa height field of the $2 \times 2.5°$ CONTROL experiment for the period 6–11 March 1993. These particular fields were extracted on the Goddard CRAY from the UNITREE files for experiments 0162 (analyses) and 0166 (forecasts) and transfered

to AER's Cambridge office for further study.

3 Methodology

In many cases the difference between two fields may be described verbally in terms of displacements and amplifications. In the present work, an amplification is treated as a bias correction for reasons to be described shortly. It is our purpose here to specify a mathematical description analogous to the verbal description, and to develop a methodology to determine that mathematical description quantitatively. The immediate difficulty is that the x- and y-displacements and the amplification or bias are themselves fields. Moreover we require that these distortion fields vary at least as smoothly as the original fields. Thus a spectral representation is appropriate. Determining the distortion which provides the best match is then equivalent to minimizing the misfit between the first field and a distortion of the second, with respect to the spectral coefficients of the distortion. In addition we impose side constraints to insure the distortion found is reasonable.

A detailed discussion of our technical approach and several examples demonstrating the utility of our approach are presented in Hoffman *et al.* (1995). In that work the distortion was assumed to be constant over the horizontal domain—that is, the distortion was completely specified by a single number for each component of the distortion (the x–displacement, the y–displacement, and the amplification). More recently, Hoffman and Grassotti (1996) extended this approach by using a double sine series representation for each component of the distortion.

In the present project we use a global or hemispheric domain, and spherical harmonics. Additionally, we treat the amplification as a bias correction for ease of interpretation. A bias correction is used instead of an amplification as a convenience. In this preliminary work, rather than apply the amplification to the difference between the field and a climate mean field, we would apply the amplification to the full field. But the effect of the amplification is then dominated by the global average of the field times the amplification, which is equivalently modeled as a slowly varying additive or bias correction. The bias correction is more straightforward to interpret.

In brief, the distortion is determined by minimizing the objective function J, by varying the displacement and bias correction fields, where

$$J = J_r + J_d + J_a.$$

Since the distortion is represented spectrally, it is the spectral coefficients of the fields of displacement and bias correction which are the independent variables. The residual cost function, J_r, measures the misfit of the distorted forecast to the verifying analysis. Minimizing J_r improves the agreement between the (distorted) forecast and the analysis. The two additional penalty terms in the objective function, J_d and J_a, ensure that the final distortion produced by the minimization is relatively smooth and not too large. (The terms cost function, objective function and penalty function are used more or less interchangeably in the literature. Here, the objective function is the quantity to be minimized, a cost function measures lack of fit to data and a penalty function measures lack of fit to a constraint.) The smoothness penalty function, J_d measures the roughness of the x- and y-displacements

and of the bias correction, ensuring that the distortion is large scale. The barrier penalty function, J_a, measures the magnitude of the distortion components in a way so that small distortions are not penalized, but large distortions are penalized heavily. This has the effect of setting up a barrier to the size of the distortions which are determined. These last two terms are evaluated using the spectral coefficients of the distortion.

For the work to date, the spectral truncations used are so severe that J_d and J_a are found to be unnecessary. These penalty functions are not used in the results presented here, except as noted. For completeness, the three terms making up J are described in the following sections. Note that the limits used to define J_a are used to precondition the minimization in most of our experiments.

3.1 Residual cost function, J_r

The residual cost function J_r measures the misfit between the distorted forecast and the verifying analysis. We denote the forecast by F, the distorted forecast by P, and the verifying analysis, or what is considered truth, by T. The cost function is evaluated over the global domain via

$$J_r = \frac{\int_\sigma (P-T)^2 d\sigma}{\int_\sigma d\sigma},$$

where the integral is the surface integral over the global domain. The distorted forecast P is obtained from the unmodified forecast F by adding a location-dependent bias correction $B(\lambda,\theta)$ to the values displaced by the displacement vector field $\boldsymbol{D}(\lambda,\theta) = (D_u, D_v)$, where $\boldsymbol{D}$ is expressed here in terms of its zonal and meridional components, in analogy to a wind field. Thus, we may write

$$P(\lambda,\theta) = F(\lambda',\theta') + B(\lambda,\theta),$$

where the location (λ',θ') is found by following the displacement vector $\boldsymbol{D}(\lambda,\theta)$ back from its endpoint (λ,θ).

We represent the scalar field B by a truncated series of spherical harmonics, and the vector field $\boldsymbol{D}$ in terms of the spectral coefficients of the corresponding vorticity (ζ) and divergence (δ) fields. A degree of smoothness can thus easily be imposed by the truncation of the series, and further constraints can separately be imposed on the divergent and rotational parts of the displacement field. The control vector C for the optimization problem is thus composed of the spectral coefficients for B, ζ and δ:

$$C = (B, \zeta, \delta)^T.$$

Both the forecast F and the verifying analysis T are available on regular latitude-longitude grids. For evaluation of the integral, it convenient to first interpolate T to a Gaussian latitude-longitude grid, in which case the formula for J_r takes the form

$$J_r = \sum_j \frac{w_j}{N_j} \sum_i (P_{ij} - T_{ij})^2,$$

where indices i, j denote the grid point location in longitude and latitude, N_j is the number of longitude points for latitude j (this number will depend on j only for reduced Gaussian

grids), and w_j is the Gaussian weight for latitude j. These weights are normalized such that their sum over all latitudes is unity.

The first step in the evaluation of the P_{ij} requires the spectral transformation from C to B_{ij} and $(D_u, D_v)_{ij}$. The next step is the evaluation of $F(\lambda', \theta')$. Following Ritchie (1987), we define latitude-longitude points in terms of 3-dimensional cartesian vectors centered on the unit sphere. The origin point (λ', θ'), corresponding to location vector $\boldsymbol{r}$, is then found in the plane of the endpoint location vector ($\boldsymbol{g}$, corresponding to gridpoint (λ_i, θ_j)), and the displacement vector $\boldsymbol{d}$ ($\boldsymbol{d}$ is the cartesian vector corresponding to $(D_u, D_v)_{ij}$):

$$\boldsymbol{r} = \alpha \boldsymbol{g} + \beta \boldsymbol{d},$$

where the coefficients α and β are chosen to satisfy the constraint that $\boldsymbol{r}$ must lie on the surface of the sphere, and that the length of the displacement vector $\boldsymbol{d}$ is equal to the great circle distance (GCD) between $\boldsymbol{g}$ and $\boldsymbol{r}$. (In Ritchie (1987), $\boldsymbol{r}$ is determined as the projection of $\boldsymbol{g} - \boldsymbol{d}$ onto the surface of the sphere.) Since the GCD is simply the angle between $\boldsymbol{g}$ and $\boldsymbol{r}$, this results in

$$\cos |\boldsymbol{d}| = \boldsymbol{g} \cdot \boldsymbol{r} = \alpha \boldsymbol{g} \cdot \boldsymbol{g} + \beta \boldsymbol{g} \cdot \boldsymbol{d} = \alpha,$$

because $\boldsymbol{g} \cdot \boldsymbol{g} = 1$ and $\boldsymbol{d}$ is orthogonal to $\boldsymbol{g}$. The remaining coefficient is obtained by multiplying the equation for $\boldsymbol{r}$ by $\boldsymbol{r}$ and using the identity $\boldsymbol{r} \cdot \boldsymbol{r} = 1$, which results in

$$|\boldsymbol{d}|\beta = -\sqrt{1 - \alpha^2},$$

where the negative sign is chosen because $\boldsymbol{d}$ is subtracted from $\boldsymbol{g}$.

In applying the above formalism, the vector $\boldsymbol{g}$ at point (i, j) is obtained from the geometric relations

$$\begin{aligned} g_x &= \cos \lambda_i \cos \theta_j, \\ g_y &= \sin \lambda_i \cos \theta_j, \\ g_z &= \sin \theta_j, \end{aligned}$$

and the coordinates (λ', θ') are obtained from $\boldsymbol{r}$ using the inverse relationships

$$\begin{aligned} \lambda' &= \tan^{-1} \frac{r_y}{r_x}, \\ \theta' &= \sin^{-1} r_z. \end{aligned}$$

(We actually use `atan2` to remove the ambiguity of the $\tan^{-1}$ function). In this coordinate system the North Pole is $(0, 0, 1)^T$ and the equator at Greenwich is $(1, 0, 0)^T$. The vector $\boldsymbol{d}$ can be obtained from (D_u, D_v) via:

$$\begin{aligned} d_x &= -D_u \sin \lambda_i - D_v \cos \lambda_i \sin \theta_j, \\ d_y &= D_u \cos \lambda_i - D_v \sin \lambda_i \sin \theta_j, \\ d_z &= D_v \cos \theta_j, \end{aligned}$$

or it can be obtained by the relation

$$\boldsymbol{d} = D_u \hat{\boldsymbol{e}}_u + D_v \hat{\boldsymbol{e}}_n,$$

where the $\hat{e}_u$ and $\hat{e}_n$ are the unit vectors (in the cartesian cooordinate system) in the direction of local east and north.

Finally, the value $F(\lambda', \theta')$ must be obtained by bilinear interpolation in longitude and latitude from the surrounding grid point values. We use F on its original regular latitude-longitude grid for this interpolation, with periodic boundary conditions in the zonal direction. No special treatment of the poles is needed for bilinear interpolation, since $|\theta'| \leq \pi/2$.

3.2 Smoothness penalty function, J_d

The smoothness penalty function, J_d is given by a simple quadratic form in terms of the spectral coefficients of the distortion,

$$J_d = \Sigma_j w_j C_j^2,$$

where j ranges over the ordering of the spectral wavenumber vectors, k, and over the components of the distortion—B, ζ, δ.

To determine the w_j, we consider the part of J_d due to the bias correction component of the distortion. In continuous form this is given by,

$$J_{d,B} = \frac{1}{\sigma_B^2} \int \int (\nabla^{2\nu} B)^2.$$

Here ν is an adjustable parameter normally taken to be 1 and σ_B is the scale for B. In the baseline case it is chosen as one tenth the standard deviation of the forecast field. Larger values of ν result in greater smoothing by emphasizing the contributions of higher wavenumbers to J_d. Using the spectral representation of B, the eigenstructure $\nabla^2 {\Psi_n}^m = -n(n+1){\Psi_n}^m$ and the orthonormality of the spherical harmonics ${\Psi_n}^m$, we find that,

$$J_{d,B} = \Sigma_k w_{k,B} C_{k,B}^2,$$

with

$$w_{k,B} = (n_k(n_k+1))^{2\nu}/\sigma_B^2.$$

The $w_{k,\zeta}$ and $w_{k,\delta}$ could be defined in the same fashion, if the smoothness constraints were to be imposed on the vorticity and divergence fields. However, since these are already first derivatives of the displacement vector field, we choose instead to impose a constraint on the size of the vorticity and divergence fields themselves. This is accomplished by setting $\nu = 0$, which amounts to using a constant weight for all the spectral coefficients of vorticity and divergence:

$$w_{k,\zeta} = 1/\sigma_d^2 \quad \text{and} \quad w_{k,\delta} = 1/\sigma_d^2$$

where σ_d is the scale of the displacement in radians.

3.3 Barrier penalty function, J_a

The functional J_a serves to limit the amplitude of the distortion. For efficiency the limits are set on the spectral coefficients. In addition the spectral coefficient limits provide a good scaling (or conditioning) for the minimization. These limits are chosen in such a way that

the grid point (or physical space) values of B and displacement at all locations will not exceed $\nu_a = 3$ times the corresponding scales for bias correction and displacement. The nominal physical space scales are $\sigma_b \sim 35\ m$ for bias correction and $\sigma_d = 0.3$ (radians) for displacement. These scales as well as the multiplier ν_a are adjustable parameters. The physical space bias correction limit is thus $B_{\text{lim}} = \nu_a \sigma_b \sim 100\ m$. For eastward and northward displacement components the physical space limit is $U_{\text{lim}} = \nu_a \sigma_d / \sqrt{2} \sim 0.63$ radians (or approximately 4000 km).

The form of J_a is chosen to be,

$$J_a = \Sigma_j (C_j / S_j)^{2\mu},$$

where j again ranges over the ordering of the spectral wavenumber vectors, k, and over (B, ζ, δ), μ is an adjustable parameter normally 10, and S_j are the spectral limiting values for the C_j. The parameter μ controls the steepness of the barrier in spectral space (Fig. 1). Each term $(C_j/S_j)^{2\mu}$ is very small (large) when C_j is smaller (larger) than S_j. This has the effect of keeping C_j smaller or the same size as S_j. In other words, the barrier penalty functional provides constraints on the C_j which have some flexibility. This seems proper since we do not *a priori* have any exact constraints.

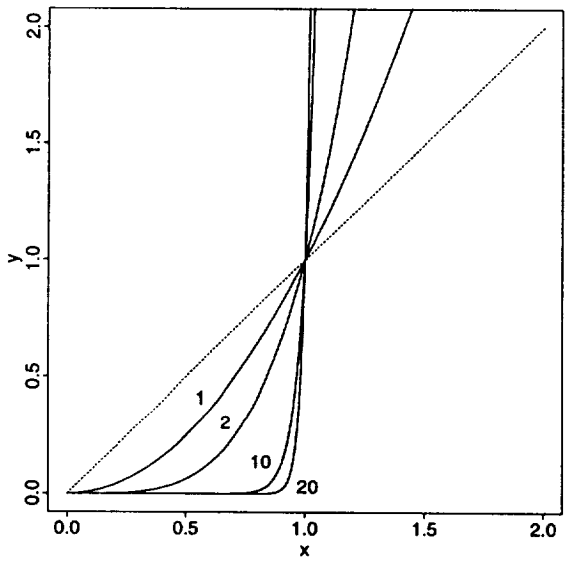

Figure 1: The component of the barrier function for a single term ($y = x^{2\mu}$ where $x = C_j/S_j$), for μ equal to 1, 2, 10, 20.

There is no unique way of setting the spectral limits. We choose limits which correspond to an equipartitioning, among the spectral modes, of the contributions to the physical space bias correction or displacement component. Here mode means each pair (m, n). The reasoning for this is that no matter what the signs of the spectral coefficients, the modes will tend to add up somewhere in the physical domain. On the other hand, the contributions within a particular mode, for example due to the sine and cosine components, are always out of phase and therefore add up in an *rms* sense, as we will see. The limits on components are chosen to correspond to a further equipartitioning.

Here, we give the derivation of the spectral limits for the bias correction which is the simpler case. The calculation for the displacement is considerably more complicated, but follows the same outline and is given in appendix A.

The bias correction, or indeed any scalar, may be represented spectrally as

$$B(\lambda,\mu) = \sum_{m=-M}^{M} \sum_{n=|m|}^{N_m} b_n^m P_n^m(\mu) e^{im\lambda}.$$

Here μ is the sine of latitude, the b_n^m are the spectral coefficients, and the P_n^m are the normalized associated Legendre functions of the first kind. Since B is real valued this may be specialized to

$$B(\lambda,\mu) = \sum_{n=0}^{N_0} \Re(b_n^0) P_n^0(\mu) + \sum_{m=1}^{M} \sum_{n=m}^{N_m} 2 P_n^m(\mu) \Re(b_n^m e^{im\lambda}).$$

We note that usually $|P_n^m| \leq 1$ and we take this to be true always for our analysis. For T10 truncation, most values of $|P_n^m|$ are smaller than 1, but close to the poles, particularly at small m, values of $|P_n^m|$ exceed 1. For $m = 0$, values of $|P_n^m|$ increase monotonically with n close to the pole, with $|P_{10}^0(1)| \sim 3.1$. Now,

$$\begin{aligned} |\Re(b_n^m e^{im\lambda})| &\leq |b_n^m e^{im\lambda}| \\ &\leq |b_n^m| = (\Re(b_n^m)^2 + \Im(b_n^m)^2)^{1/2} \\ &\leq \sqrt{2}\max(|\Re(b_n^m)|, |\Im(b_n^m)|). \end{aligned}$$

Therefore,

$$|B(\lambda,\mu)| \leq \sum_{n=0}^{N_0} |\Re(b_n^0)| + \sum_{m=1}^{M} \sum_{n=m}^{N_m} 2\sqrt{2}\max(|\Re(b_n^m)|, |\Im(b_n^m)|).$$

As a result, if N is the total number of modes, and if we choose the spectral limits as

$$\max |\Re(b_n^0)| \leq B_{\text{lim}}/N$$

and

$$\max(|\Re(b_n^m)|, |\Im(b_n^m)|) \leq B_{\text{lim}}/(2\sqrt{2}N),$$

for $m > 0$, then we must have

$$|B(\lambda,\mu)| \leq B_{\text{lim}}$$

everywhere.

3.4 Implementation details

The algorithm is implemented in Splus, except that the spectral transform (and computation of Gaussian latitudes and weights) use a set of Fortran library functions. All computations are performed in double precision. To minimize J we use the built-in Splus function `nlmin`, which implements the algorithms of Dennis *et al.* (1981) and which uses function values only (first and second derivatives of the cost function are estimated by finite differences, using repeated evaluations of the cost function). This approach is computationally inefficient, but appropriate for the prototyping and demonstration of year 1 of the project.

There are several different options that can be used:

1. Any pentagonal truncation can be used. In addition, selected spectral coefficients can be removed from the control vector and kept at a constant value (usually zero) throughout the minimization. Currently the (0,0) spectral coefficient (which is proportional to the global average) is not included for any of the control variables and is held constant at 0. In addition, this feature is used for implementing the next two options.

2. The minimization can be performed with or without bias correction.

3. A global or hemispheric domain can be used. A hemispheric domain speeds the minimization because it requires only half the degrees of freedom for a given truncation. (However, the spectral transforms and Gaussian sums are still performed over the global domain, which is symmetric about the equator).

4. The Gaussian grid used for evaluating the forecast errors can be coarser resolution than the original forecast and verification grids. The original forecast and verification grids are equally spaced 2.5° longitude by 2° latitude grids (corresponding to 144 longitude and 91 latitude points over the globe); in all but the preliminary runs we used a Gaussian grid with 5° resolution in longitude, and 46 latitude points over the globe. This resolution is sufficient for the scale of the features we are interested in, while providing a four-fold savings in computation time.

5. The penalty functions for the limiting values barrier function and smoothing function can be turned on or off. The penalty functions are turned off for all of the results presented here, except as noted.

6. The minimization can be run with different estimates of the scale of the control variables: uniform scale, scales derived from the smoothing function weights, and scales derived from the limiting values. The latter produces the best convergence results, and is used to prepare all of the results presented here, except as noted.

7. Various tuning parameters can be adjusted for the minimization.

4 Results

Results presented below show that the methodology works, that a large part of the total error may be explained by a distortion limited to $T10$ truncation (i.e., triangular truncation at wavenumber 10), and that the remaining residual error contains mostly small spatial scales. With forecasts separated by 24 h, time continuity of the distortion fields describing the forecast errors is present in some areas, but lacking in other areas.

Experiments so far are all based on the first of the CONTROL set of forecasts described in Section 2. The nominal experiment is for a northern hemispheric 72 h forecast of 500 hPa height and its corresponding verifying analysis valid at 00 UTC 9 March 1993. In the nominal case, we use $T10$ truncation and a Gaussian grid with half the resolution of the forecast fields to represent the distortion, and do not use the barrier and smoothing functions. Differences from this nominal setup will be noted below.

Preliminary real data tests using $T5$ truncation were unsatisfactory and spectral analysis of the forecast and verification fields indicated that a truncation at wavenumber 10-15 would be more appropriate. The test runs using synthetic data (Section 4.2) indicate that the algorithms have been properly implemented, but that convergence of the minimization is sensitive to the scaling of the control vector.

In this study, we use finite difference estimates of the gradient of J. Unfortunately the minimization algorithm sometimes has difficulties moving away from a zero initial estimate of the distortion. The function J_r is only piecewise differentiable, because the bilinear interpolation is not smooth at the latitude-longitude box boundaries. Experience shows that this problem can be ignored in general, because only a tiny fraction of points end up on the box boundaries. However for a zero initial estimate, all interpolation points are at the corners of the boxes, resulting in maximum non-differentiability. As a result the minimization algorithm may abnormally terminate. Therefore, we generally use random initial estimates. Once near the minimum, the minimization algorithm stops in different places for different starting locations (Section 4.5). In what follows, unless otherwise noted, the results presented correspond to the first solution found. However, the solutions found are generally all very similar. Further, the objective function in the neighborhood of the minimum is well behaved (Fig. 15), suggesting that with a direct calculation of the gradient, these difficulties should be absent.

4.1 Preliminary real data tests

Our first test used $T5$ truncation, global fields and a full resolution Gaussian grid for the distortion. The minimizations stopped after only a few iterations, explaining only a small part of the forecast error. It was apparent from visual examination of the fields, that the forecast errors are smaller scale than the distortion fields, and only the largest scales of the forecast errors are affected by the displacement and bias correction fields. We concluded that we must allow the distortion to include scales at least as large as the synoptic scales.

Later real data experiments experiments for the nominal case (using a $T10$ truncation) showed that scaling the control vector by the smoothing weights quickly leads to false convergence, with only small changes to the objective function, whereas scaling the control vector by its limiting values results in a successful minimization.

For the nominal case, turning on the barrier and smoothness penalty functions results in distortions which are smaller in magnitude, but larger in scale, and residual errors which are larger in scale and magnitude. The residual rms error is 42 m for this configuration, as compared to 26 m for the unconstrained nominal case. The spectra of the original and residual forecast error, both with and without the penalty functions (Fig. 2) show that a great deal of the forecast error on the scales of the distortion is explained by the distortion and that the penalty constraints have a strong effect on the smallest scales in the distortion. In this figure, the energy (more properly the variance) is plotted as a function of total wavenumber n. That is, the values are the sum of squares of all spectral coefficients having the same value of n.

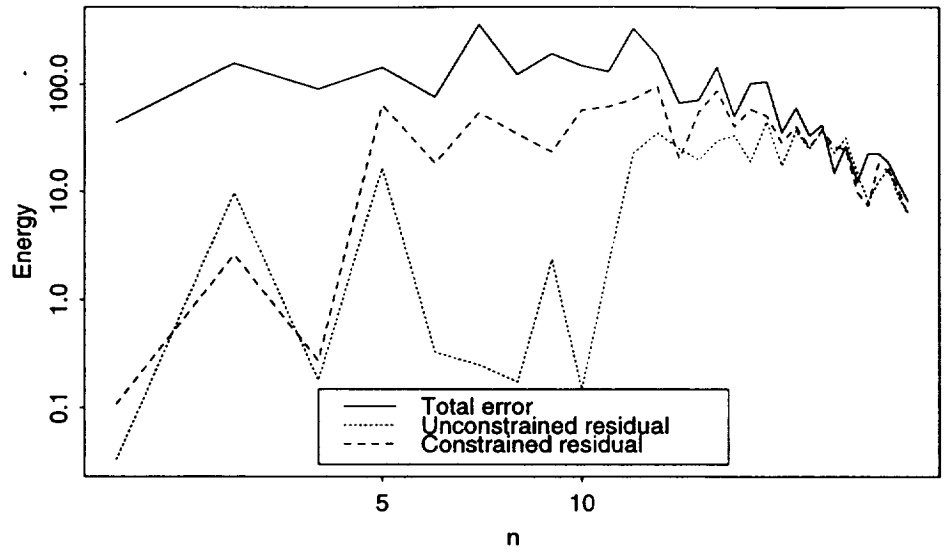

Figure 2: Total wavenumber spectrum of the original forecast error (solid curve) and of the residual error calculated from the unconstrained distortion (dotted curve) and the constrained distortion (dashed curve).

4.2 Test runs using synthetic data

The convergence of the algorithm was tested using synthetic data. A synthetic verification was obtained by distorting the 3-day 500 hPa height forecast with a known, random distortion (random numbers with a uniform distribution and a range given by the limiting values of the control vector), and using the resulting distorted field in place of the verification. A successful minimization should be able to recover the true distortion from any reasonable first guess. In this case a Gausian grid comparable in resolution to the original $2 \times 2.5°$ grid was used.

At $T5$ and $T10$ truncation, the true solution is quickly recovered for the case when no bias correction is used in either the true distortion or during the minimization, even when a uniform scaling is used for the control vector. For example, in this situation, the solid curve (Case A) in Fig. 3, is calculated for uniform scaling and a $T10$ truncation. This is also true for the case when the bias correction is turned on both during the minimization and for the true distortion (Case C in Fig. 3, short dashed curve).

If the bias correction is turned off for the true distortion, but allowed to be nonzero during the minimization, results depend critically on the proper scaling of the control vector. If the control vector is scaled by its limiting values estimate, the true solution is quickly recovered (Case B, Fig. 3, dotted curve). If the scaling is derived from the smoothing function instead, the minimization quickly fails with false convergence. For the case of uniform scaling of the control vector, the minimization is only partially successful (Case D, Fig. 3, long dashed curve): the objective function is reduced only slowly, and after 100 iterations, only half of the original forecast error is explained.

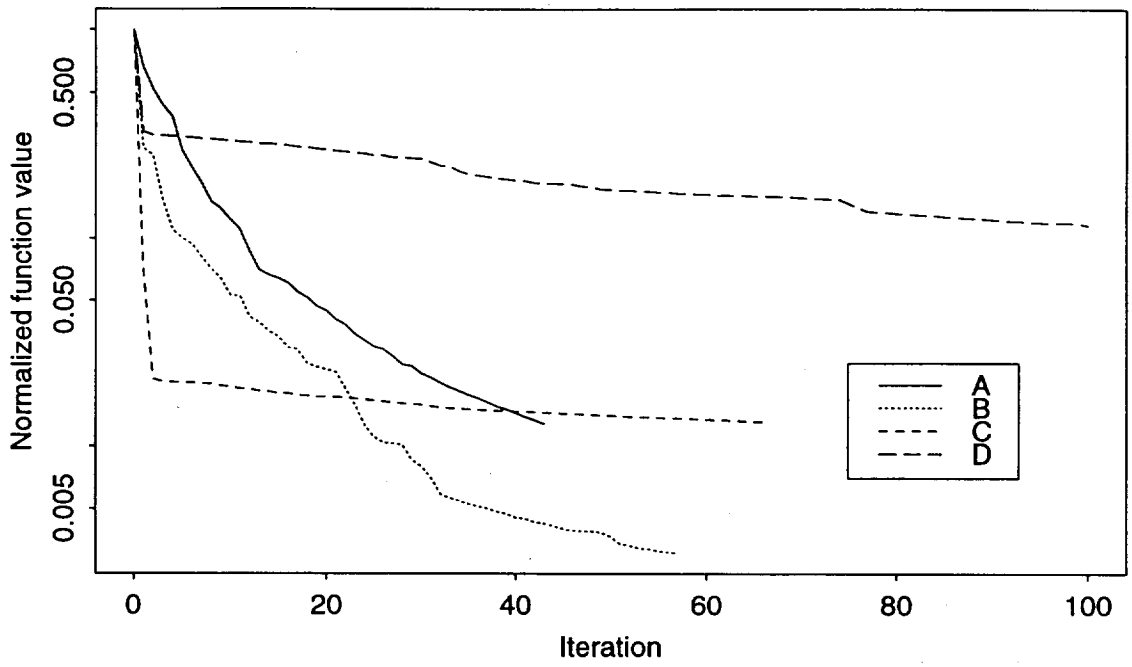

Figure 3: The value of the objective function, normalized by its initial value, as a function of iteration number for the cases discussed in the text.

4.3 Results for the nominal case

These tests used the three-day forecast (Fig. 4) and its verifying analysis (Fig. 5) of 500 hPa height, evaluated over the northern hemisphere on a Gaussian grid with half the resolution of the original forecast and verification grids. All tests use a $T10$ truncation.

For the nominal case, a solution is reached after 17 iterations with a reduction in the rms residual error from 64 m to 26 m, which corresponds to a six-fold reduction in the objective function. The distortion field has significant displacements (exceeding 500 km in some locations) and bias corrections (20 m).

Visual examination of the distortion field reveals a number of features with phase and amplitude errors. At the Alaska coast, the original forecast has a single short-wave to the north, which is deeper than in the verifying analysis, and a hint of two troughs to the south. The verifying analysis shows a second shortwave to west. The error field (Fig. 6) shows a dipole of negative errors at the Alaska coast, and positive errors to the south and west. The distortion field (Fig. 7) partially corrects these errors by northwestward displacements, and a negative/positive bias correction dipole pattern. The resulting distorted forecast (Fig. 8) is qualitatively much closer to the analysis, as is also apparent in the plot of the residual forecast error (Fig. 9).

Another prominent feature in the original error field is over Scotland, where the ridge is under-forecast and the low to the north is over-forecast. The distortion field accounts for this with northward displacements over the British Isles, and increasingly eastward displacements

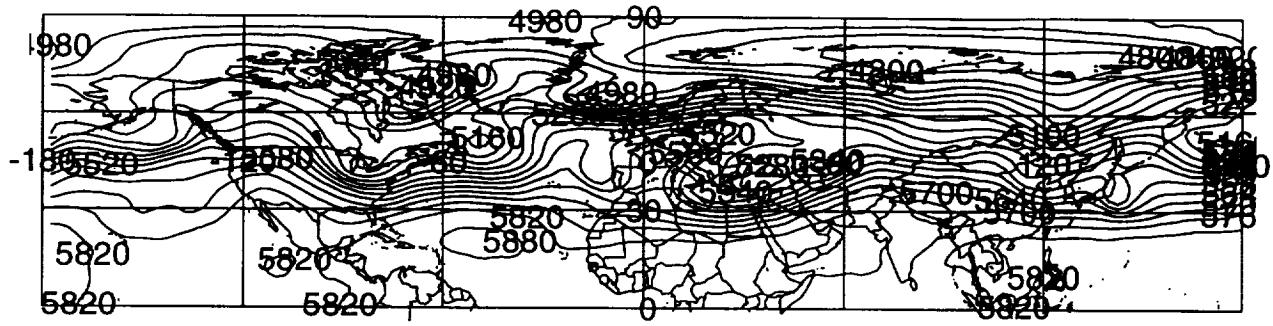

Figure 4: Unmodified forecast field for the 3-day forecast of 500 hPa height valid 00 UTC 9 March 1993.

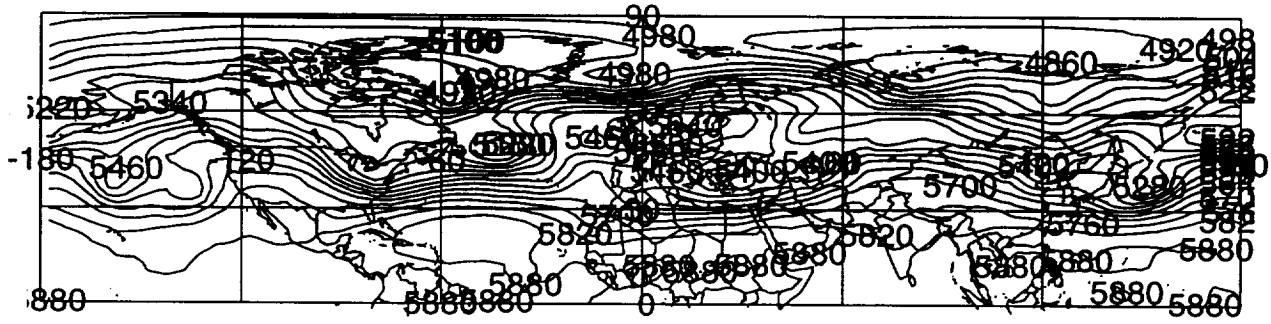

Figure 5: Verifying 500 hPa height field for 00 UTC 9 March 1993.

to the north and east. The resulting distorted forecast over that region has much improved strengths and positions of the ridge over the British Isles and the low to its north. A clear example of a phase error of a short wave is at 150°E, where the forecast position of the shortwave is too far west and south. The distortion field corrects this through northeastward displacements in combination with a broad area of negative bias correction. However, this particular feature is of such small scale that it can be explained only partially (from over 250 m to ca. 170 m). In general, the distortion fields appear to properly account for the forecast errors. Also, as desired, the bias corrections and displacements determined are small in the tropics, where the 500 hPa height fields are devoid of features and forecast errors are small.

For the day 3 forecast, we also calculated a distortion with displacements only. That is, the bias was held at zero. Naturally with fewer degrees of freedom the distortion explains less of the forecast error. The rms residual error for this experiment is 34 m, as opposed 26 m for the nominal case. This result is similar to what was found in the preliminary results for the case when the penalty functions were activated (see Fig. 2 and discussion).

4.4 Time evolution of the distortion

The time continuity of the distortion from day to day is good in some areas and involves large changes in other areas. In general, both the displacements and biases increase with forecast length. This is obvious in both the plots (to be discussed shortly) and in statistics of the solutions. Table 1 shows that the size of the forecast errors increases roughly linearly

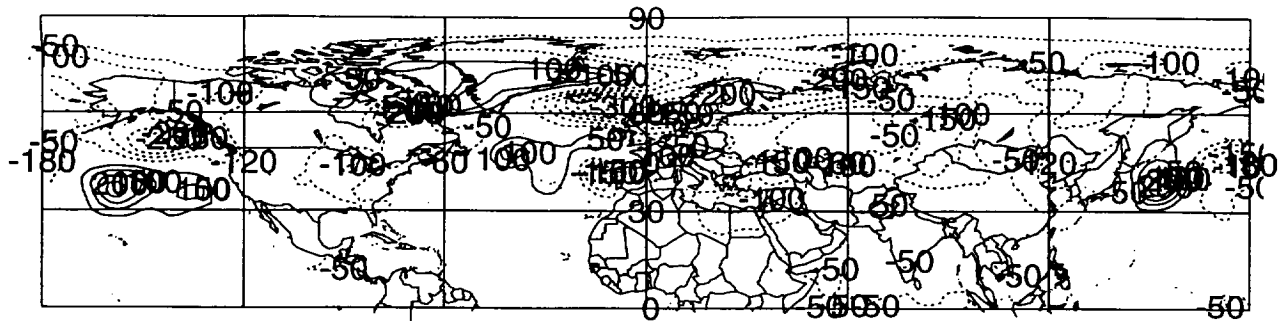

Figure 6: Original forecast errors for the 3-day forecast of 500 hPa height valid 00 UTC 9 March 1993.

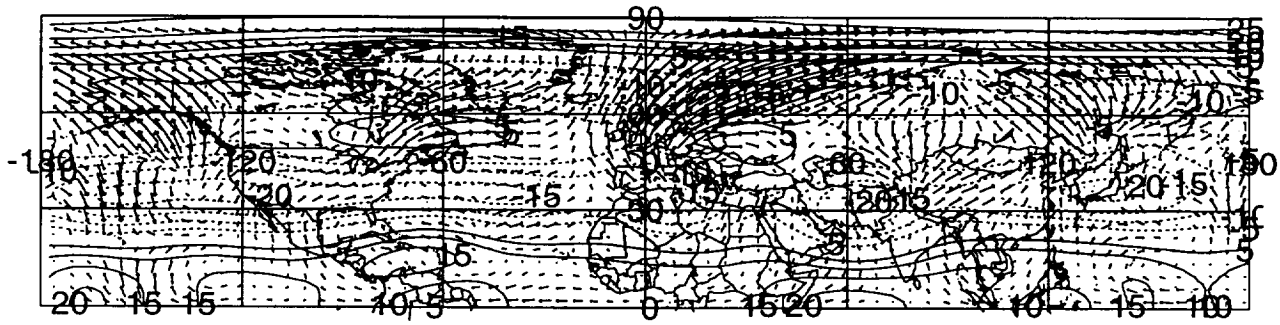

Figure 7: Distortion field for the nominal case. Displacement vectors are shown over a contour plot of the bias correction.

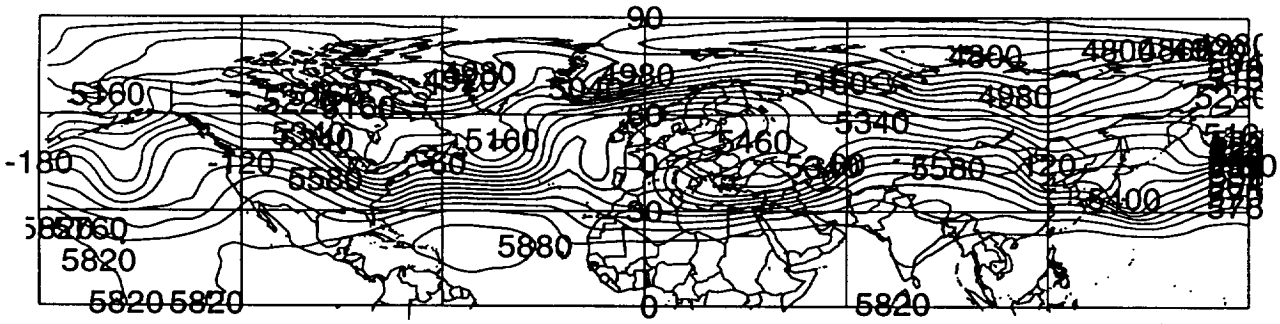

Figure 8: Distorted forecast for the nominal case.

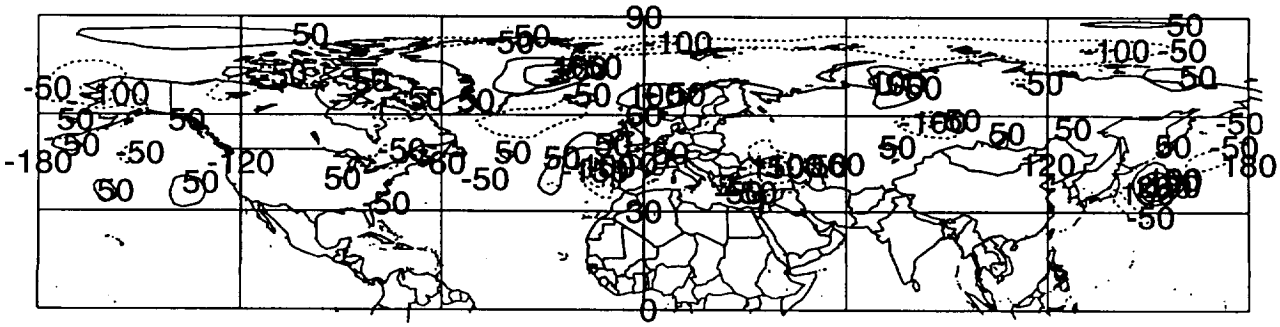

Figure 9: Residual forecast error for the nominal case.

with time and that the distortion explains approximately half of the forecast error standard deviation at all times.

In Table 1 the statistics are calculated as follows for each run for the distortion corresponding to the initial estimate and to the solution. The rms error is the rms residual error for the distortion (in m). The rms gradient is the rms scaled gradient (in m^2), calculated as

$$(\sum_j (S_j \frac{\partial J}{\partial C_j})^2)^{1/2},$$

and the distortion size is the dimensionless distance in scaled phase space, calculated as

$$(\sum_j (C_j/S_j)^2)^{1/2}.$$

The distortion size may equivalently be calculated as $J_a^{1/2}$ with $\mu = 1$. Finally, in Table 1, N is the number of iterations used by the minimization.

Table 1: Summary statistics for forecast days 1–5, with day 3 being the nominal case. The calculation of the statistics is described in the text.

	Initial RMS		Distortion	Final RMS		Distortion	
Day	Error	Gradient	Size	Error	Gradient	Size	N
1	31.20	506	3.85	17.26	72	6.65	14
2	45.28	746	3.83	23.52	121	8.53	12
3	61.47	1087	3.86	29.80	116	10.75	10
4	73.18	1186	3.91	33.85	225	14.07	9
5	85.41	1268	3.99	39.96	159	15.92	8

Figs. 10–14 show the evolution of the distortion and error components for days 1–5, for an area in the central North Pacific. The distortion bias in this area has good time continuity. The overall pattern of negative bias in the belt near 40°N, with centers of negative bias at both edges of the domain and a smaller negative or positive bias at the north and south edges of the domain, is relatively constant over the 5 forecast days. The displacement field is considerably more variable in time. In some areas the displacement vectors change significantly from map to map, while in other areas there is a more consistent evolution. The error components generally have poor time continuity. This is especially so for the residual errors. Some features seem to persist between pairs of maps. However, since it is hard to trace any error feature to a third map, it seems unlikely that this persistence is significant.

There are several potential reasons for the lack of consistency in the evolution of the distortion. One possibility is the use of random initial estimates for the minimization. However as described in Section 4.5, it does not appear that the overall solutions are sensitive to this. Another possibility is that the absence of penalty functions in the current work leads to overfitting. Penalizing the smallest scales may eliminate some of the temporal inconsistencyin the sequence of the distortions if this is the cause. Another possibility is that 24 h is too long a time increment to expect such temporal consistency.

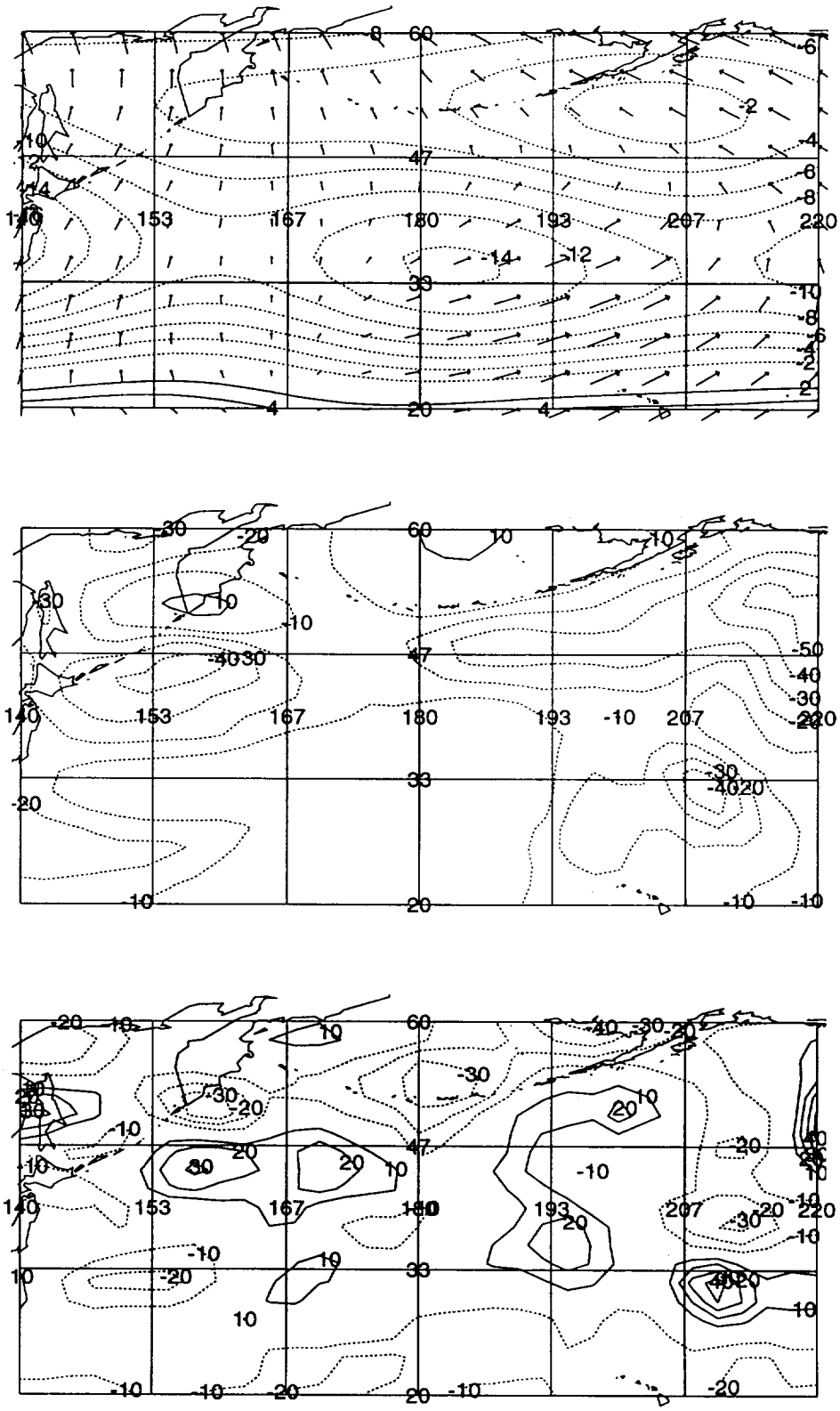

Figure 10: The distortion (top), distortion error (middle) and residual error (lower) in the North Pacific for the one day forecast valid 00 UTC 7 March 1993.

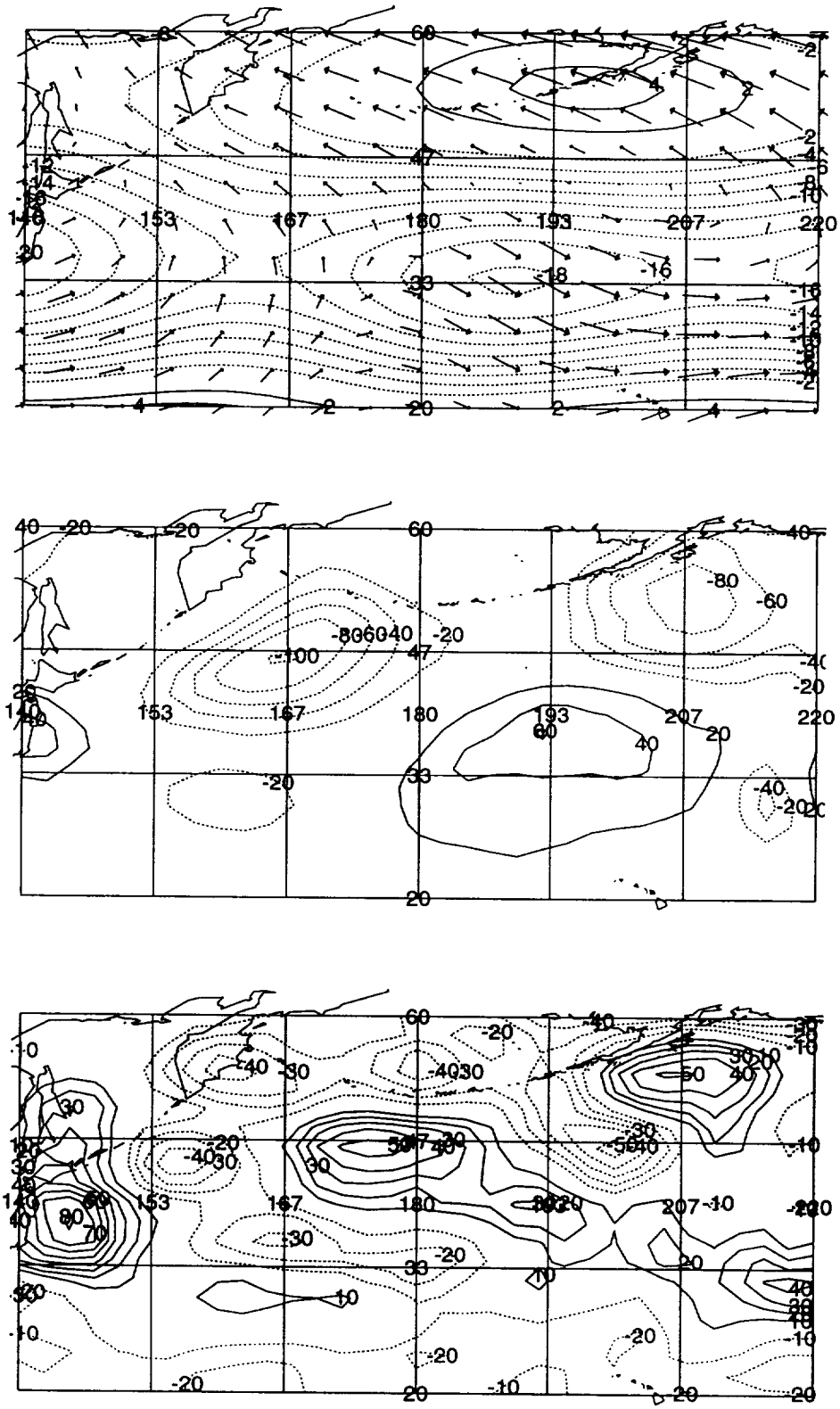

Figure 11: The distortion (top), distortion error (middle) and residual error (lower) in the North Pacific for the two day forecast valid 00 UTC 8 March 1993.

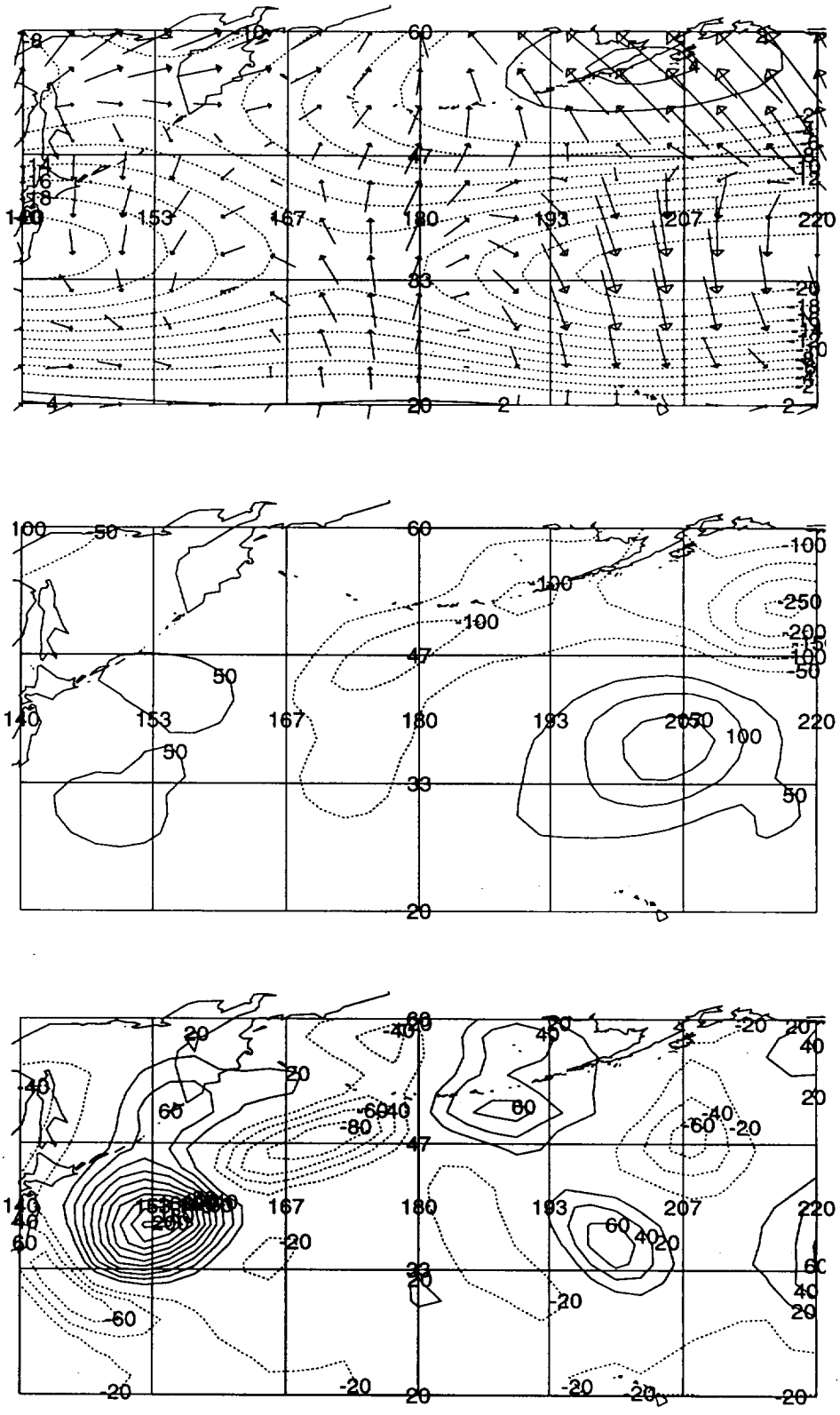

Figure 12: The distortion (top), distortion error (middle) and residual error (lower) in the North Pacific for the three day forecast valid 00 UTC 9 March 1993.

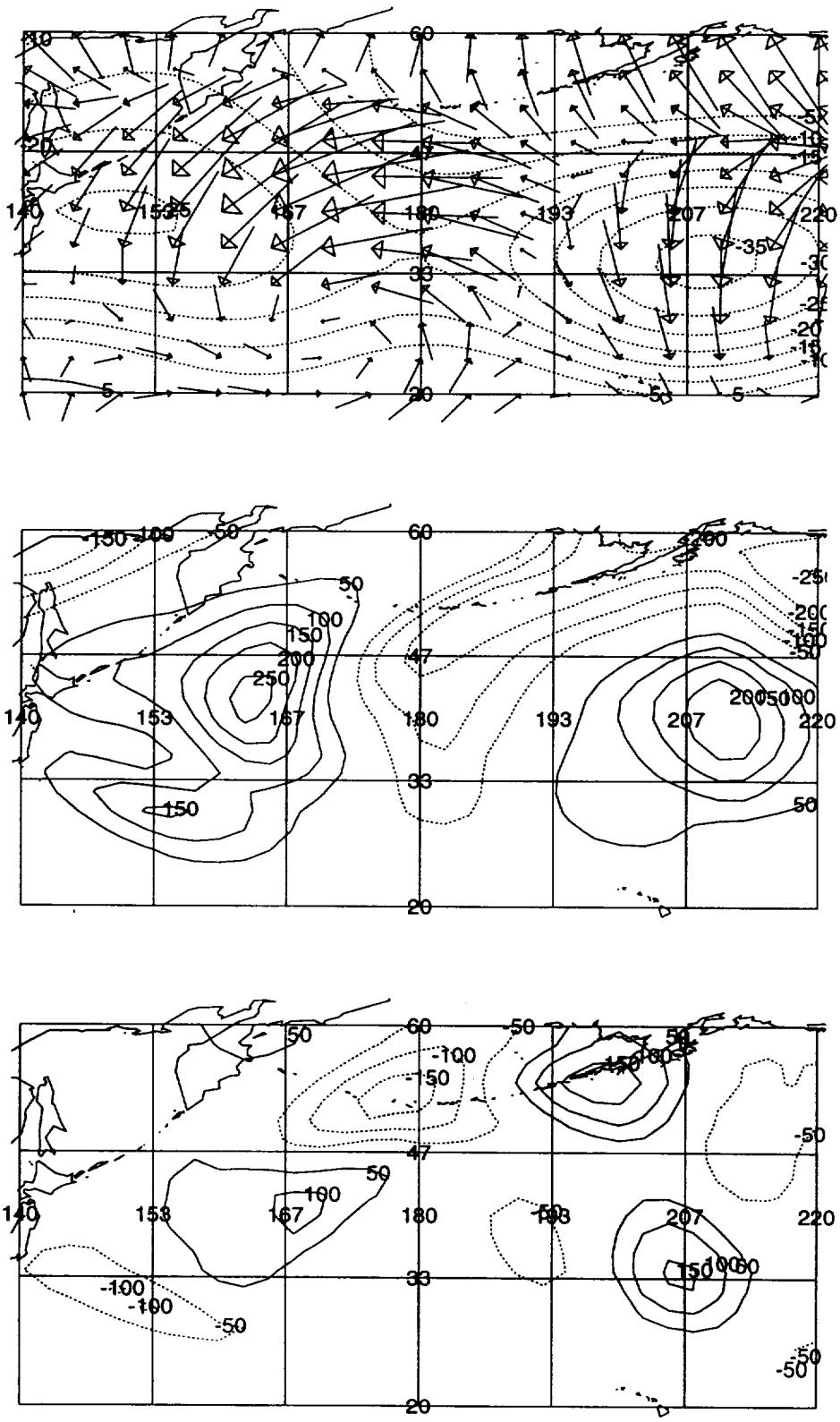

Figure 13: The distortion (top), distortion error (middle) and residual error (lower) in the North Pacific for the four day forecast valid 00 UTC 10 March 1993.

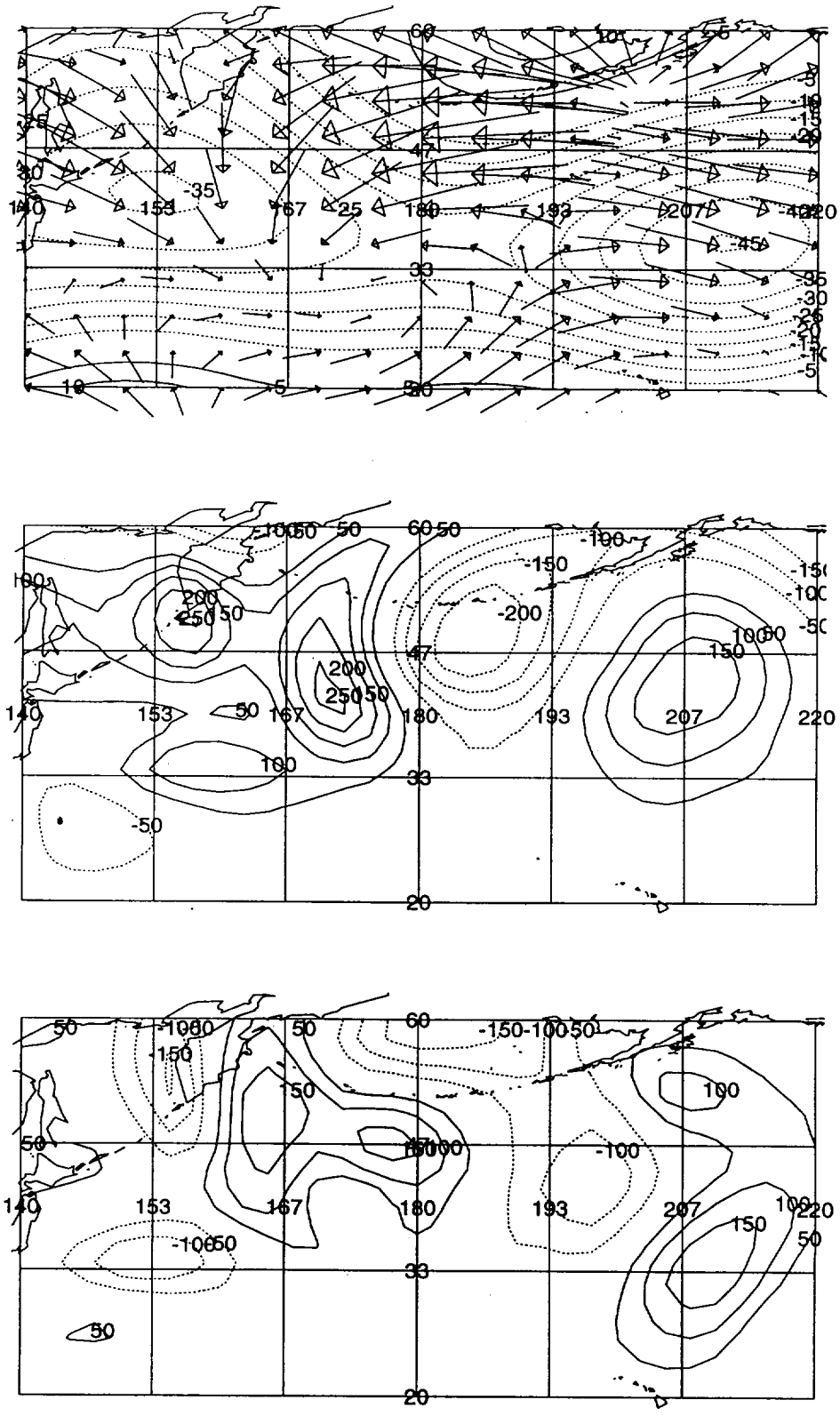

Figure 14: The distortion (top), distortion error (middle) and residual error (lower) in the North Pacific for the five day forecast valid 00 UTC 11 March 1993.

4.5 Reproducibility

The minimization sometimes terminates immediately with a "false convergence" condition when starting from the reasonable initial estimate of zero distortion. As discussed, this is a point of maximum non-differentiability and the finite difference calculation of the gradient here is very sensitive to the step size. To avoid this problem, we set the initial estimate to random numbers, uniform on [-1,1] scaled by half the limiting value for each corresponding entry in the control vector.

Use of random initial estimates brings with it a concern for reproducibility. To gauge the effect of the random initial estimates we repeated some of the runs up to five times. In nearly all cases, the final results are close together, both in terms of the value of the objective function and the rms scaled gradient (see Table 2), and in terms of the visual appearance of the distortion. Differences between solutions are probably a result of insufficient convergence due to the use of a finite difference gradient. In addition, the largest value of the objective function in the ensemble of solutions is associated with fewest number of iteration, suggesting that the minimization did not advance sufficiently.

The greatest difference between solutions occurred in the experiment not allowing any bias in the distortion. In this case the largest and smallest rms residual errors were 40 m and 34 m respectively. Fig. 15 shows the value of the objective function along the line defined by these two solutions. Along the line, parameterized in terms of s, the absolute minimum falls at $s \sim -0.5$, while the two solutions found correspond to $s = 0$ and $s = 1$. The smooth parabolic shape of the curve in Fig. 15 suggests that the objective function is very well behaved in the vicinity of the true minimum and with the use of a direct calculation of the gradient, the difficulties described here will cease.

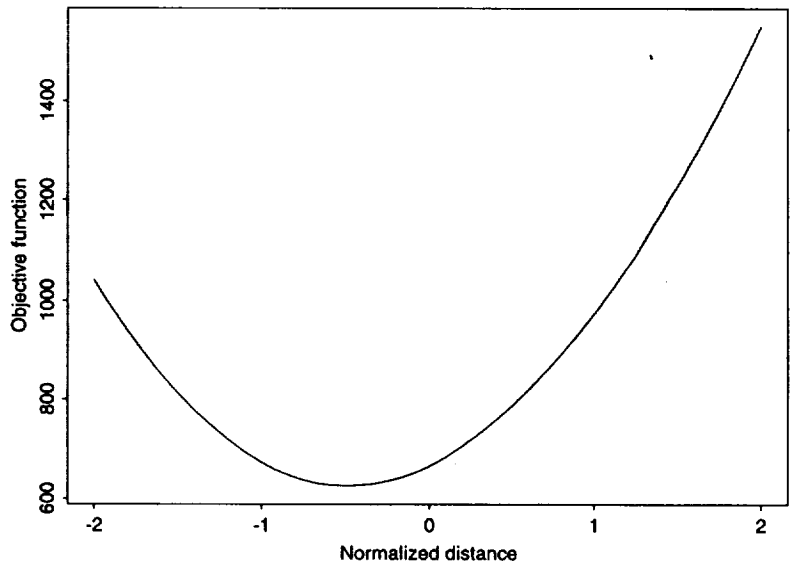

Figure 15: Value of the objective function J along the line in phase space between two numerical solutions. The line is parameterized in terms of s, with the two solutions at $s = 0$ and $s = 1$.

Table 2: Summary statistics for the nominal case for an ensemble of initial estimates For comparison the value of the rms error for a zero initial estimate is 61.16 m. The calculation of the statistics is described in the text in the discussion of Table 1.

	Initial RMS		Distortion	Final RMS		Distortion	
Run	Error	Gradient	Size	Error	Gradient	Size	N
1	61.47	1087	3.86	29.80	116	10.75	10
2	63.25	1170	3.92	27.52	141	13.12	12
3	60.51	1040	3.89	31.19	132	10.21	7
4	60.34	1030	3.94	30.28	87	10.63	8
5	63.89	1097	3.98	25.82	103	14.98	17

5 Plans for future work

Our immediate plans to extend the current research fall into several related subtopics, as described in the following paragraphs.

5.1 Other forecast variables: SLP

The forecasts of SLP for the case of 6 March 1993 for $24, 48, 72, 96, 120$ h will be analyzed. Other fields are possible. Differences in the distortion fields for different forecast fields will be interesting.

5.2 Other cases: ERS 1 data impact

The analyses can be extended to other cases in the suite of experiments. So far we have used only the Control run for 6 March. We have access to cases for 6, 11, 16 and 21 March for Control, ESA, PNMC, PGLA and VARGLA experiments.

A complete evaluation of the errors of all these experiments would constitute a major study. However at least one other forecast time and one of the experiments using ERS 1 data will be analyzed.

5.3 Initialization experiments

Experiments will test the suitability of extrapolating distortions in time to provide initial estimates for the minimization. For example, the initial estimate of the 72 h distortion might be taken to be $1.5C_{48\ h}$ or $2C_{48\ h} - C_{24\ h}$.

This may improve efficiency. More importantly, it may improve continuity in the time evolution of the distortion.

5.4 Parameter sensitivity studies

We will examine the sensitivity of the distortion and error fields to the parameters of the method $\nu, \sigma_d, \sigma_A, \nu_a, \mu$.

The use of different spectral truncations and grid resolutions will be studied.

Different ways of setting the spectral limits and/or scales will be explored. In particular, we will try to eliminate some of the assumptions in the derivation of the spectral limits by using the actual maximum amplitudes of the expansion functions evaluated on the transform grid.

5.5 Computation efficiency

Fortran codes and corresponding adjoints will be developed for J to directly calculate the gradient, thereby allowing the use of more efficient minimization routines.

References

[1] R. Atlas, R. N. Hoffman, E. Brin, and P. M. Woiceshyn. The impact of ERS-1 scatterometer data on GEOS model forecasts. In *International Symposium on Assimilation of Observations in Meteorology and Oceanography*, Tokyo, Japan, 13-17 Mar. 1995. WMO.

[2] J. E. Dennis, D. M. Gay, and R. E. Welsch. An adaptive nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software*, 7:348–383, 1981.

[3] R. N. Hoffman and C. Grassotti. A technique for assimilating SSM/I observations of marine atmospheric storms. *Journal of Applied Meteorology*, 35(8):1177–1188, Aug. 1996.

[4] R. N. Hoffman, Z. Liu, J.-F. Louis, and C. Grassotti. Distortion representation of forecast errors. *Monthly Weather Review*, 123(9):2758–2770, Sept. 1995.

[5] H. Ritchie. Semi-Lagrangian advection on a Gaussian grid. *Monthly Weather Review*, 115:608–619, 1987.

[6] S. D. Schubert, R. B. Rood, and J. Pfaendtner. An assimilated dataset for earth science applications. *Bulletin of the American Meteorological Society*, 74(12):2331–2342, Dec. 1993.

A Limiting values for displacement spectral coefficients

In section 3.3 we calculated the limiting values for B_n^m, the spectral coefficients for the bias correction. The goal here is to derive similar limiting values for the eastward displacement component, D_u. The derivation for the northward displacement component, D_v, is entirely analogous. In what follows, recall that the form of J_a was chosen for convenience and because there is no exact physical constraint on the size of the distortion, and that the parameters U_{lim} may be considered adjustable. Consequently in setting the spectral limits, we are free to make some approximations and to choose simpler but looser bounds.

The eastward displacement component is obtained from the pseudo-displacement $D_U = \cos(\phi)D_u$, which in turn is represented spectrally as

$$D_U(\lambda,\mu) = \sum_{m=-M}^{M} \sum_{n=|m|}^{N_m} K_n[\zeta_n^m H_n^m(\mu) + im\delta_n^m P_n^m(\mu)]e^{im\lambda}.$$

Here

$$H_n^m = -\cos(\phi)\frac{dP_n^m}{d\phi},$$

and for convenience we have defined the spectral Laplacian operator L_n and its inverse K_n by

$$L_n = K_n^{-1} = -n(n+1).$$

In the following derivation, we have neglected the factor $1/\cos(\phi)$ and replaced D_U by D_u throughout. This amounts to imposing looser limits at high latitudes than over the rest of the domain: for the coarse Gaussian grid used in most experiments desribed here, limits on the actual displacement are less than two (four) times those of the pseudo-displacements for all but the seven (three) northernmost latitudes in each hemisphere.

Since D_u is real valued its representation may be specialized to

$$D_u(\lambda,\mu) = \sum_{n=1}^{N_0}[K_n H_n^0(\mu)]\Re(\zeta_n^0) + \sum_{m=1}^{M}\sum_{n=m}^{N_m}[K_n H_n^m(\mu)]2\Re(\zeta_n^m e^{im\lambda}) + [mK_n P_n^m(\mu)]2\Re(i\delta_n^m e^{im\lambda}).$$

The terms in square brackets are all smaller in magnitude than $1/n$. This follows from the definition of K_n, the assumption that $|P_n^m| \le 1$, and the fact that $m \le n$ for each mode in the truncation, provided that $|H_n^m| < n+1$. This last bound is established below.

Now if all the terms in square brackets are $< 1/n$, and following the derivation of Section 3.3, we see that

$$|D_u(\lambda,\mu)| < \sum_{n=1}^{N_0}\frac{1}{n}\Re(\zeta_n^0) + \sum_{m=1}^{M}\sum_{n=m}^{N_m}\frac{4}{n}\max(|\Re(\zeta_n^m)|, |\Im(\zeta_n^m)|, |\Re(\delta_n^m)|, |\Im(\delta_n^m)|).$$

Here we make the approximation that P_n^m and H_n^m are out of phase in μ in the same way as sine and cosine. Therefore the four terms—$H_n^m(\mu)\cos(m\lambda)\Re(\zeta_n^m)$, $-H_n^m(\mu)\sin(m\lambda)\Im(\zeta_n^m)$, $-mP_n^m(\mu)\cos(m\lambda)\Im(\delta_n^m)$, $-mP_n^m(\mu)\sin(m\lambda)\Re(\delta_n^m)$—can add up at most in an *rms* sense. It follows that if we choose the spectral limits as

$$\max(|\Re(\zeta_n^0)|, |\Re(\delta_n^0)|) \le nU_{\text{lim}}/N$$

and

$$\max(|\Re(\zeta_n^m)|, |\Im(\zeta_n^m)|, |\Re(\delta_n^m)|, |\Im(\delta_n^m)|) \leq nU_{\lim}/(4N),$$

for $m > 0$, then we must have

$$|D_u(\lambda, \mu)|, |D_v(\lambda, \mu)| \leq U_{\lim}$$

everywhere.

It remains to show that $|H_n^m| < n + 1$. By definition

$$H_n^m = n\varepsilon_{n+1}^m P_{n+1}^m - (n+1)\varepsilon_n^m P_{n-1}^m$$

where

$$\varepsilon_n^m = \frac{1}{2}\left(\frac{n^2 - m^2}{n^2 - (\frac{1}{2})^2}\right)^{\frac{1}{2}}.$$

For $m > 0$, $n^2 - m^2 < n^2 - (1/2)^2$, so $\varepsilon_n^m < 1/2$. For $m = 0$, $\varepsilon_n^m > 1/2$ but approaches $1/2$ as n increases.

Now

$$|H_n^m| \leq n\varepsilon_{n+1}^m + (n+1)\varepsilon_n^m \equiv h_n^m,$$

assuming again that $|P_n^m| \leq 1$. For $m > 0$ we then have

$$|H_n^m| < n + \frac{1}{2} < n + 1.$$

For $m = 0$, dropping the superscript 0, for $n = 1, 2, 3, \ldots, \infty$, we directly calculate that $\varepsilon_n = 0.577, 0.516, 0.507, \ldots, 0.5$, and that $h_n = 1.671, 2.563, 3.540, \ldots, n + 1/2$. This establishes that $|H_n^m| < n + 1$ for $m = 0$ as well, provided only that $dh_n/dn < 1$.

To show $dh_n/dn < 1$, note that

$$d\log\varepsilon_n = (1 - 4\varepsilon_n^2)d\log n.$$

Therefore,

$$\frac{dh_n}{dn} - 1 = -g(\varepsilon_{n+1}) - g(\varepsilon_n) - \frac{4}{n}\varepsilon_n^3 < 0,$$

because $n > 1$ and $\varepsilon_n > 1/2$. Here

$$g(x) \equiv 4x^3 - 2x + 1/2 > 0$$

for all $x > 1/2$ since $g(1/2) = 0$ and $g'(x) = 12x^2 - 2 > 0$ for all $x > 1/2$.

The spectral limits could be made more precise by actually calculating the maximum absolute values of P_n^m and H_n^m and using these values and the value of $\cos(\phi)$ to define the spectral limits of the actual, not pseudo, displacements. This may be pursued in the future, but for the work so far and especially considering the low spectral truncations used, the limits derived here have proven adequate.

B Required forms

The required Report Documentation Pages (NASA Form 1626 and Standard Form 298) are attached.

NASA National Aeronautics and Space Administration

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle
Distortion Representation of Forecast Errors for Model Skil Assessment and Objective Analysis

5. Report Date
8/8/96

6. Performing Organization Code

7. Author(s)
Ross N. Hoffman, Thomas Nehrkorn and Christopher Grassotti

8. Performing Organization Report No.
P599

10. Work Unit No.

9. Performing Organization Name and Address
Atmospheric and Environmental Research, Inc.
840 Memorial Drive
Cambridge, MA 02139

11. Contract or Grant No.
NAS5-32953

13. Type of Report and Period Covered
Tech. Report 8/8/96

12. Sponsoring Agency Name and Address
NASA/GSFC
Earth Sciences Procurement Office
Greenbelt, MD 20771

14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

We study a novel characterization of errors for numerical weather predictions. In its simplest form we decompose the error into a part attributable to phase errors and a remainder. The phase error is represented in the same fashion as a velocity field and will be required to vary slowly and smoothly with position. A general distortion representation allows for the displacement and a bias correction of forecast anomalies. In brief, the distortion is determined by minimizing the objective function by varying the displacement and bias correction fields. In the present project we use a global or hemispheric domain, and spherical harmonics to represent these fields.

In this project we are initially focusing on the assessment application, restricted to a realistic but univariate 2--dimensional situation. Specifically we study the forecast errors of the 500 hPa geopotential height field for forecasts of the short and medium range. The forecasts are those of the Goddard Earth Observing System data assimilation system.

Results presented show that the methodology works, that a large part of the total error may be explained by a distortion limited to triangular truncation at wavenumber 10, and that the remaining residual error contains mostly small spatial scales.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Numerical Weather Prediction Forecast Errors	Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	25	

NASA FORM 1628 OCT 86

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 8, 1996	Technical Report 8/8/96

4. TITLE AND SUBTITLE
Distortion Representation of Forecast Errors for Model Skill Assessment and Objective Analysis

5. FUNDING NUMBERS
NAS5-32953

6. AUTHOR(S)
Ross N. Hoffman, Thomas Nehrkorn and Christopher Grassotti

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Atmospheric and Environmental Research, Inc.
840 Memorial Drive
Cambridge, MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER
P599

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
NASA/GSFC
Earth Sciences Procurement Office
Greenbelt, MD 20771

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

We study a novel characterization of errors for numerical weather predictions. In its simplest form we decompose the error into a part attributable to phase errors and a remainder. The phase error is represented in the same fashion as a velocity field and will be required to vary slowly and smoothly with position. A general distortion representation allows for the displacement and a bias correction of forecast anomalies. In brief, the distortion is determined by minimizing the objective function by varying the displacement and bias correction fields. In the present project we use a global or hemispheric domain, and spherical harmonics to represent these fields.

In this project we are initially focusing on the assessment application, restricted to a realistic but univariate 2--dimensional situation. Specifically we study the forecast errors of the 500 hPa geopotential height field for forecasts of the short and medium range. The forecasts are those of the Goddard Earth Observing System data assimilation system.

Results presented show that the methodology works, that a large part of the total error may be explained by a distortion limited to triangular truncation at wavenumber 10, and that the remaining residual error contains mostly small spatial scales.

14. SUBJECT TERMS
numerical weather prediction
forecast errors

15. NUMBER OF PAGES
25

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited